**Breast Cancer Prediction**

End Term Report

*By*

Aman Pandey And Aditya Raj Jha

Section: KM118

Roll Nos.: RKM118B51 And RKM118A17



Department of Intelligent Systems

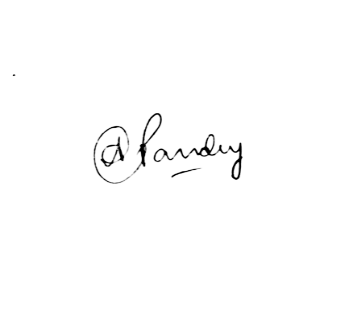
School of Computer Science Engineering

Lovely Professional University, Jalandhar

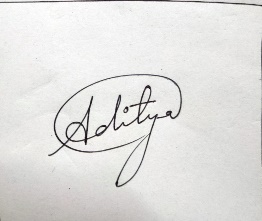
November – 2022

**Student Declaration**

This is to declare that this report has been written by me/us. No part of the report is copied from other sources. All information included from other sources have been duly acknowledged. I/We aver that if any part of the report is found to be copied, I/we are shall take full responsibility for it.



Aman Pandey

RKM118B51

Aditya Raj Jha

RKM118A17

Lovely Professional University

Jalandhar

05/11/2022

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| Title | Page No. |
| Introduction | 5 |
| Magazines | 3 |
| Notebooks | 1 |
| Paper pads | 1 |
| Pens | 3 |
| Pencils | 2 |
| Highlighter | 2 colors |
| Scissors | 1 pair  6 |

**BONAFIDE CERTIFICATE**

Certified that this project report “Breast Cancer Prediction” is the bonafide work of “Aman Pandey AND Aditya Raj Jha” who carried out the project work under my supervision

Dr. Dhanpratap Singh

Associate professor

25706

Intelligence System 1

**Introduction About the Topic**

Breast cancer has become the most recurrent type of health issue among women especially for women in middle age. Early detection of breast cancer can help women cure this disease and death rate can be reduced . In the present-day scenario, to observe breast cancer mammograms are used and they are known be the most effective scanning technique. In this paper the detection of cancer cells is done by machine learning technique.

Image processing is a method to convert an image into digital form and perform some operations on it, to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image . Usually, Image Processing system includes treating images as two-dimensional signals while applying already set signal processing methods to them. Image processing basically includes the following three steps: • Importing the image with optical scanner or by digital photography. • Analyzing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs. • Output is the last stage in which result can be altered image or report that is based on image analysis.

Digital Processing techniques help in manipulation of the digital images by using computers. Digital Image is composed of a finite number of elements, each of which elements have a particular value at a particular location. These elements are referred to as picture elements, image elements and pixels. A Pixel is most widely used to denote the elements of a Digital Image. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing.

The three general phases that all types of data have to undergo while using digital technique are Preprocessing, enhancement and display, information extraction. The first step is to analyses the images and represent them. The method of image representation will fix fundamental problems that are the ineffectiveness of capturing textural information and the weak capacity for classification of features that result in low performance of retrieval. In Content-based Image Retrieval, similarity estimation is a part of the primary task and has a greater effect on the accuracy of retrieval and time of retrieval. The project aims to solve problems such as "Which similarity measure is appropriate for particular feature type and how to reduce the similarity calculation computation? "And" The texture function is more representative and discriminatory to describe the mammogram of the given query? In this project we use classification techniques such as Fuzzy C-means and KNN to employ feature extraction.

The FCM clustering is used for image segmentation once the mammographic image was collected. Data point in this system be held by to various clusters with changing degrees of membership and is based on objective criteria. The segmented area is rigorously examined using Multi-level Discrete Wavelet Transform to get edge details to which is then used as a feature. PCA is then used for this data to analyze along with GLCM. After performing the analysis, 13 features extracted in the proposed framework and their pixel values in matrix form are stored in database.

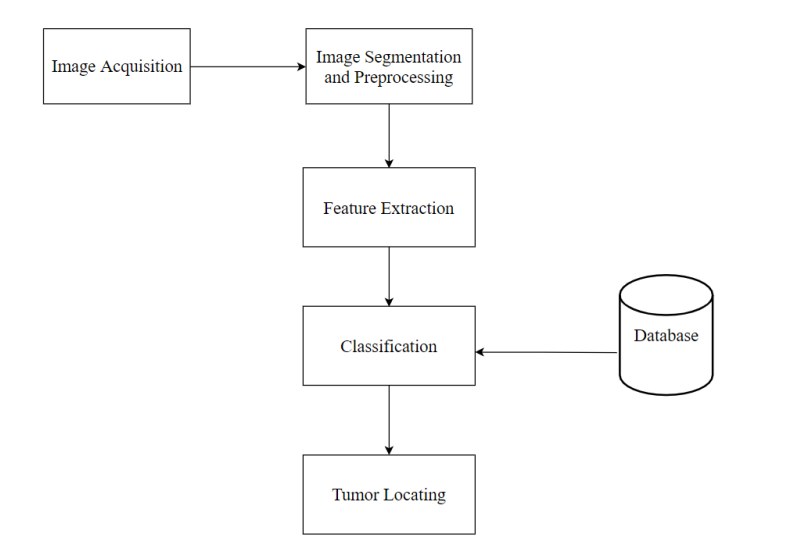
Machine learning is an application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. The process of learning begins with observations or data, such as examples, direct experience, or instruction, to look for patterns in data and make better decisions in the future based on the

examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly

Supervised algorithms with machine learning skills to provide both input and desired output, in addition to furnishing feedback about the accuracy of predictions during algorithm training. A classification problem is when the output variable is a category or a group. Using the KNN classifier method, which primarily relies on the shape of the cancer cells in the image, the algorithm classifies the image into Benign, Malignant and Normal once the features are extracted & fully educated. By performing appropriate morphological operations, the device calculates the appropriate region properties, such as size, Euler number, etc., and displays the detected boundary image along with the tumor area.

**Method Of Working**

The figure 1 shows system architecture of breast cancer prediction and tracking system. The system mainly consists of four processes. Once the image is acquitted, the system converts the image into gray scale image. By applying the suitable Image segmentation techniques, the system aims at extracting a meaningful object lying in the image. Clustering is one the powerful image segmentation technique involves grouping of data points to cluster, the system implements clustering technique using Fuzzy C-means (FCM) . The segmented region is completely analyzed by using the Multi-level Discrete Wavelet Transform, Principal Component Analysis (PCA) along with Gray Level Co-occurrence Matrix (GLCM) features . Totally 13 features are extracted in the system and their pixel values in the form of matrix is stored in database that is in the db. mat file, some of the features extracted by the system are mean, variance, entropy etc. Then image undergoes classification process with respect to dataset in db. mat file and classifies the image into Benign, Malignant and Normal. The system also performs morphological operations and calculates region properties of the image such as Area, Eccentricity and Euler number. If the image has cancer cells, then the tumor area is computed and displayed by the system along with the boundary detected image.



(pictorial flow of project)